## Part 1: Design, Models, Perception

Information Visualization Mini-Course
TECS Week 2008

Tamara Munzner

UBC Computer Science

7 January 2008



## Information Visualization

- visual representation of abstract data
  - computer-generated, often interactive



## Interactivity

- static images
  - 10,000 years
  - art, graphic design
- moving images
  - 100 years
  - cinematography
- interactive graphics
  - 20 years
  - computer graphics, human-computer interaction



## Information Visualization

- visual representation of abstract data
  - computer-generated, often interactive
  - help human perform some task more effectively



## Information Visualization

- visual representation of abstract data
  - computer-generated, can be interactive
  - help human perform some task more effectively
- bridging many fields
  - graphics: drawing in realtime
  - cognitive psych: finding appropriate representation
  - HCI: using task to guide design and evaluation
- external representation
  - reduces load on working memory
  - offload cognition
  - familiar example: multiplication/division



## External Representation: multiplication

| paper | mental buffer |
|---|---|
| 57 | |
| x 48 | |
| —– | |



## External Representation: multiplication

| paper | mental buffer |
|---|---|
| 5 | |
| 57 | |
| x 48 | [ 7*8=56] |
| —– | |
| 6 | |



## External Representation: multiplication

| paper | mental buffer |
|---|---|
| 5 | |
| 57 | |
| x 48 | [5*8=40 + 5 = 45] |
| —– | |
| 456 | |



## External Representation: multiplication

| paper | mental buffer |
|---|---|
| 2 | |
| 57 | |
| x 48 | [7*4=28] |
| —– | |
| 456 | |
| 8 | |



## External Representation: multiplication

| paper | mental buffer |
|---|---|
| 2 | |
| 57 | |
| x 48 | [5*4=20+2=22] |
| —– | |
| 456 | |
| 228 | |



## External Representation: multiplication

| paper | mental buffer |
|---|---|
| 57 | |
| x 48 | |
| —– | |
| 1 | |
| 456 | |
| 228 | |
| —— | |
| 36 | [8 + 5 = 13] |



## External Representation: multiplication

| paper | mental buffer |
|---|---|
| 57 | |
| x 48 | |
| —– | |
| 456 | |
| 228 | |
| —— | |
| 2736 | |



## Information Visualization

- visual representation of abstract data
  - computer-generated, can be interactive
  - help human perform some task more effectively
- bridging many fields
  - graphics: drawing in realtime
  - cognitive psych: finding appropriate representation
  - HCI: using task to guide design and evaluation
- external representation
  - reduces load on working memory
  - offload cognition
  - familiar example: multiplication/division
  - infovis example: topic graphs



## External Representation: Topic Graphs

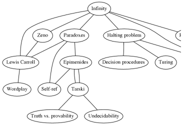
[Godel, Escher, Bach. Hofstadter 1979]

Turing - Halting problem
Halting problem - Infinity
Paradoxes - Lewis Carroll
Paradoxes - Infinity
Infinity - Lewis Carroll
Infinity - Unpredictably long searches
Infinity - Recursion
Infinity - Zeno
Infinity - Paradoxes
Lewis Carroll - Zeno
Lewis Carroll - Wordplay
Halting problem - Decision procedures
BlooP and FlooP - AI
Halting problem - Unpredictably long searches
BlooP and FlooP - Unpredictably long searches
BlooP and FlooP - Recursion
Tarski - Truth vs. provability
Tarski - Epimenides
Tarski - Undecidability
Paradoxes - Self-ref
...



## External Representation: Topic Graphs

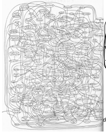
- offload cognition to visual systems
- minimal attention to read answer





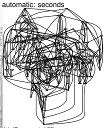

## External Rep: Automatic Layout

manual: hours, days



(Godel, Escher, Bach. Hofstader 79)

automatic: seconds



dot, (Gansner et al 93)

## Mini-Course Outline

- Part 1: Monday morning
  - Intro
  - Design Studies
  - Models
  - Perception
- Part 2: Monday afternoon
  - Color
  - Space, Layers, and Ordering
  - Statistical Graphics
- Part 3: Thursday afternoon
  - Multiples and Interaction
  - Navigation and Zooming
  - Focus+Context
- Part 4: Friday morning
  - High Dimensional Data
  - Graphs and Trees
  - User Studies



## Design Studies

- two concrete infovis examples
  - before presenting theoretical models
- design study: infovis solution for specific application
  - carry out task analysis in target domain
  - derive design requirements
  - justify choices of visual encoding and interaction
  - build prototype
  - (refine until satisfied)



## Cluster-Calendar, van Wijk

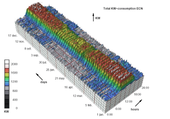
van Wijk and van Selow, Cluster and Calender based Visualization of Time Series Data, Proc. InfoVis 99, http://www.win.tue.nl/~vanwijk/clv.pdf

- data: N (value, time) pairs
  - N large: 50K
- tasks
  - find standard day patterns
  - find how patterns distributed over year, week, season
  - find outliers from standard daily patterns
  - want overview first, then detail on demand
- possibilities
  - predictive mathematical models
    - details lost, multiscale not addressed
  - scale-space approaches (wavelet, fourier, fractal)
    - hard to interpret, known scales lost
  - 3D extruded mountain: x hours, y value, z days
    - 3D often pretty but not useful



## 3D Problems

- occlusion, perspective distortion
- seasonal visible, but daily/weekly hard to see

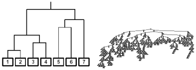



[van Wijk and van Selow, Cluster and Calender based Visualization of Time Series Data, InfoVis99, http://www.win.tue.nl/~vanwijk/clv.pdf]



## Create Clusters

- create derived variable by transforming raw data
- use day as fundamental unit based on task analysis
- hierarchical clustering, merging most similar days
- but not interesting to show dendrogram directly
  - structure of dendrogram hierarchy does not address requirements

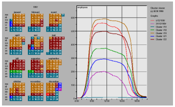



[van Wijk and van Selow, Cluster and Calender based Visualization of Time Series Data, InfoVis99, http://www.win.tue.nl/~vanwijk/clv.pdf]



## Link Curves and Calendar

- show clusters as linked curves+calendars in 2D
- see patterns
  - office hours, fridays in/and summer, school break
  - weekend/holidays, post-holiday, santa claus



[van Wijk and van Selow, Cluster and Calender based Visualization of Time Series Data, InfoVis99, http://www.win.tue.nl/~vanwijk/clv.pdf]



## Lessons

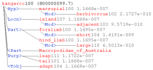
- derived variable: clusters
- challenges of naive 3D usage
- linked displays
- traditional visual representation of time: calendar
- clear task analysis guided choices
  - reject standard 3D extrusion
  - reject standard dendrogram
- critique
  - colors not maximally discriminable



## Linguistic Networks, Munzner

Munzner, Guimbretiere, and Robertson. Constellation: A Visualization Tool For Linguistic Queries from MindNet. Proc. InfoVis 99. http://graphics.stanford.edu/papers/const

- data: MindNet query results
- definition graph from dictionary entry sentence
  - nodes: word senses
  - links: relation types






## Semantic Network

- definition graphs used as building blocks
- unify shared words
- large network
  - millions of nodes
  - grammar checking now, translation future
  - global structure known: dense
- probes return local info



## Path Query

- best N paths between two words
- words on path itself

kangaroo100—Part→forelimb100—Mod→short104—Join→short←Mod—tail100

- definition graphs used in computation

```
kangaroo100   (vole101 tapir100 s
sharp-tailed_grouse100 scut100 r
pitta100 partridge104 lynx100 lo
kingfisher100 horned_toad100 haw
bobtail101 bobtail100 bobcat100
Scottish_terrier100)
```

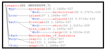


## Task: Plausibility Checking

- paths ordered by computed plausibility
- researcher hand-checks results
  - high-ranking paths believable?
  - believable paths high-ranked?
  - are stop words all filtered out?



## Window Flipping Problem





## Goal

- create a unified view of relationships between paths and definition graphs
  - shared words are key
  - thousands of words (not millions)
- special purpose algorithm debugging tools
  - not understand structure of English

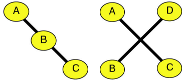


## Constellation Video



## Traditional Graph Layout

- avoid crossings using careful spatial positioning
- reason: avoid false attachments

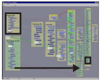



[Munzner, Interactive Visualization of Large Graphs and Networks (PhD thesis), Stanford University, 2000, http://graphics.stanford.edu/papers/munzner_thesis]



## Information Visualization Approach

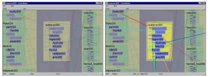
- encode domain-specific attribute with spatial position
  - why? spatial position is strongest perceptual cue
  - plausibility gradient from left to right
- novel layout algorithm using curvilinear grid
  - paths as backbone, attach definition graphs
  - many crossings for long-distance proxy links



[http://graphics.stanford.edu/papers/munzner_thesis]

## Visual Layering

- avoid perception of attachments with visual layers
- interactively highlight sets of boxes and edges
  - perceptual channels: size, saturation, brightness
- avoid hidden state of visible/invisible toggle
  - implicit assumption: if not visible, doesn't exist
  - previous control actions often forgotten



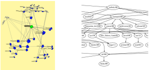
[http://graphics.stanford.edu/papers/munzner_thesis]



## Task-Oriented Design Success

- task-specific methods

- generic approaches



[http://graphics.stanford.edu/papers/munzner_thesis]



## Lessons

- spatial position as strongest perceptual cue
- interactive visual layering, avoiding hidden state
- custom application vs. generic toolkits
- critique
  - effort of custom application vs. generic toolkits



## Design Studies: Reading

Cluster and Calendar based Visualization of Time Series Data.
Jarke J. van Wijk and Edward R. van Selow
Proc. InfoVis 99, pp 4-9
http://www.win.tue.nl/~vanwijk/clv.pdf

Constellation: A Visualization Tool For Linguistic Queries from MindNet.
Tamara Munzner, François Guimbretière, and George Robertson.
Proc. InfoVis 99, p 132-135
http://graphics.stanford.edu/papers/const

Constellation: Linguistic Semantic Networks
Tamara Munzner
Interactive Visualization of Large Graphs and Networks (PhD thesis) Chapter 5, Stanford University, 2000, pp 87-122
http://graphics.stanford.edu/papers/munzner_thesis



## Design Studies: Further Reading

Process and Pitfalls in Writing Information Visualization Research Papers.
Tamara Munzner, book chapter to appear.
http://www.cs.ubc.ca/~tmm/courses/infovis/readings/pitfalls.pdf

Exploratory visualization of array-based comparative genomic hybridization.
Robert Kincaid, Amir Ben-Dor, and Zohar Yakhini
Information Visualization 3(4):176–190, 2005, Palgrave Macmillan

ThemeRiver: In Search of Trends, Patterns, and Relationships.
Susan Havre, Beth Hetzler, and Lucy Nowell.
Proc. InfoVis 2000, pp 115–123

Vizster: Visualizing Online Social Networks.
Jeffrey Heer and danah boyd.
Proc. InfoVis 2005, pp 32–39
http://jheer.org/publications/2005-Vizster-InfoVis.pdf

Visual Analysis of Historic Hotel Visitation Patterns.
Chris Weaver, David Fyfe, Anthony Robinson, Deryck W. Holdsworth, Donna J. Peuquet and Alan M. MacEachren.
Information Visualization 6(1):89–103, 2007.
http://www.geovista.psu.edu/publications/2006/Weaver_vast2006.pdf

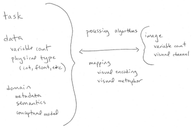


## Mini-Course Outline

- Part 1: Monday morning
  - Intro
  - Design Studies
  - Models
  - Perception
- Part 2: Monday afternoon
  - Color
  - Space, Layers, and Ordering
  - Statistical Graphics
- Part 3: Thursday afternoon
  - Multiples and Interaction
  - Navigation and Zooming
  - Focus+Context
- Part 4: Friday morning
  - High Dimensional Data
  - Graphs and Trees
  - User Studies



## Visualization Big Picture





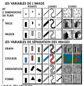

## Mapping

- input
  - data semantics
  - use domain knowledge
- output
  - visual encoding
    - visual/graphical/perceptual/retinal
    - channels/attributes/dimensions/variables
  - use human perception
- processing
  - algorithms
  - handle computational constraints



## Bertin: Semiology of Graphics

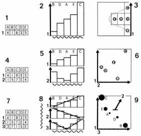
- geometric primitives: marks
  - points, lines, areas, volumes
- attributes: visual/retinal variables
  - parameters control mark appearance
  - separable channels flowing from retina to brain
- x,y
  - position
- z
  - size
  - greyscale
  - color
  - texture
  - orientation
  - shape




[Bertin, Semiology of Graphics, 1967 Gauthier-Villars, 1998 EHESS]



## Design Space = Visual Metaphors



[Bertin, Semiology of Graphics, 1967 Gauthier-Villars, 1998 EHESS]



## Data Types

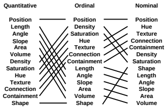
- continuous (quantitative)
  - 10 inches, 17 inches, 23 inches
- ordered (ordinal)
  - small, medium, large
  - days: Sun, Mon, Tue, ...
- categorical (nominal)
  - apples, oranges, bananas


[graphics.stanford.edu/papers/polaris]



## Channel Ranking Varies by Data Type

- spatial position best for all types




[Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986]



## Mackinlay/Card Model

- data variables
  - 1D, 2D, 3D, 4D, 5D, etc
- data types
  - nominal, ordered, quantitative
- marks
  - point, line, area, surface, volume
  - geometric primitives
- retinal properties
  - size, brightness, color, texture, orientation, shape...
  - parameters that control the appearance of geometric primitives
  - separable channels of information flowing from retina to brain

[Card, Mackinlay, and Shneiderman. Chapter 1, Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann 1999.]



## Shneiderman's Data+Tasks Taxonomy

- data
  - 1D, 2D, 3D, temporal, nD, trees, networks
  - (text and documents - Hanrahan)
- tasks
  - overview, zoom, filter, details-on-demand,
  - relate, history, extract
- data alone not enough
  - what do you need to do?
- mantra: overview first, zoom and filter, details on demand

[Shneiderman, The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations]



## Amar/Eagan/Stasko Task Taxonomy

- low-level tasks
  - retrieve value, filter, compute derived value,
  - find extremum, sort, determine range,
  - characterize distribution, find anomalies,
  - cluster, correlate
- standardized set for better comparison between papers

[Amar, Eagan, and Stasko. Low-Level Components of Analytic Activity in Information Visualization. Proc. InfoVis 05]

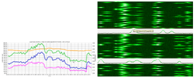


## Control Room Example

Which location has the highest power surge for the given time period? (find extreme y-dimension)

A fault occurred at the beginning of this recording, and resulted in a temporary power surge. Which location is affected the earliest? (find extreme x-dimension)

Which location has the most number of power surges? (extreme count)



[Overview Use in Multiple Visual Information Resolution Interfaces. Lam, Munzner, and Kincaid. Proc. InfoVis 2007]

## Data Types and Conceptual Models

- from raw data model
  - 17, 25, -4, 28.6
  - (floats)
- to conceptual model of semantics
  - (temperature)
- consider task
  - making toast
  - classifying showers
  - finding anamolies in local weather patterns
- to determine data type
  - burned vs. not burned (N)
  - hot, warm, cold (O)
  - continuous to 4 sig figures (Q)

[Rethinking Visualization: A High-Level Taxonomy. Melanie Tory and Torsten Möller, Proc. InfoVis 2004, pp. 151-158.]



## Combinatorics of Encodings

- challenge
  - pick the best encoding from exponential number of possibilities $(n+1)^8$
  - if using n visual channels
- Principle of Consistency
  - properties of the image should match properties of data
- Principle of Importance Ordering
  - encode most important information in most effective way



## Automatic Design

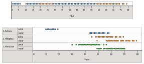
Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986

- APT system: pick encoding automatically given data
  - limited set of encodings: scatterplots, bar/line charts
- Expressiveness Criterion
  - Set of facts expressible in visual language if sentences (visualizations) in language express all facts in data, and only facts in data.
- consider the failure cases...



## Cannot Express the Facts

- A 1 ⇔ N relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position



[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]



## Expresses Facts Not in the Data

- length interpreted as quantitative value
  - Thus length says something untrue about nominal data

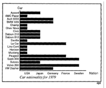



Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986]



## Automatic Design

Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986

- APT system: pick encoding automatically given data
  - limited set of encodings: scatterplots, bar/line charts
- Expressiveness Criterion
  - Set of facts expressible in visual language if sentences (visualizations) in language express all facts in data, and only facts in data.
- Effectiveness
  - A visualization is more effective than another visualization if information conveyed by one visualization is more readily perceived than information in other.
  - subject of the next section



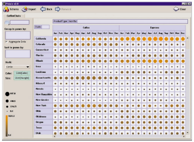
## Polaris/Tableau

Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002. http://graphics.stanford.edu/papers/polaris_extended

- infovis spreadsheet
  - table cell graphical elements, not just numbers
- user drag and drop exploration of marks/channels
  - instead of automatic selection
- table algebra ⇔ interactive interface ops
  - formal language
- commercialized via `www.tableausoftware.com`



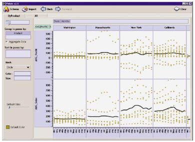
## Polaris: Circles, State/Product:Month



[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]



## Polaris: Circles, Profit/State:Months



[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]



## Models: Reading

Semiology of Graphics. Jacques Bertin, Gauthier-Villars 1967, EHESS 1998

Automating the Design of Graphical Presentations of Relational Information. Jock Mackinlay, ACM Transaction on Graphics, vol. 5, no. 2, April 1986, pp. 110-141.

Chapter 1, Readings in Information Visualization: Using Vision to Think. Stuart Card, Jock Mackinlay, and Ben Shneiderman, Morgan Kaufmann 1999.

The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations Ben Shneiderman, Proc. 1996 IEEE Visual Languages, also Maryland HCIL TR 96-13 [citeseer.ist.psu.edu/shneiderman96eyes.html]

Low-Level Components of Analytic Activity in Information Visualization. Robert Amar, James Eagan, and John Stasko. Proc. InfoVis 05 [www.cc.gatech.edu/~john.stasko/papers/infovis05.pdf]

Rethinking Visualization: A High-Level Taxonomy. Melanie Tory and Torsten Möller, Proc. InfoVis 2004, pp. 151-158. [webhome.cs.uvic.ca/ mtory/publications/infovis04.pdf]

Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG 8(1), January 2002. [graphics.stanford.edu/papers/polaris]



## Models: Further Reading

The Grammar of Graphics. Leland Wilkinson, Springer-Verlag 1999

The Structure of the Information Visualization Design Space Stuart Card and Jock Mackinlay, Proc. InfoVis 97 [citeseer.ist.psu.edu/card96structure.html]

A Function-Based Data Model for Visualization. Lloyd Treinish, Visualization 1999 Late Breaking Hot Topics

Multiscale Visualization Using Data Cubes. Chris Stolte, Diane Tang and Pat Hanrahan, Proc. InfoVis 2002



## Mini-Course Outline

- Part 1: Monday morning
  - Intro
  - Design Studies
  - Models
  - Perception
- Part 2: Monday afternoon
  - Color
  - Space, Layers, and Ordering
  - Statistical Graphics
- Part 3: Thursday afternoon
  - Multiples and Interaction
  - Navigation and Zooming
  - Focus+Context
- Part 4: Friday morning
  - High Dimensional Data
  - Graphs and Trees
  - User Studies



## Human Perception

- sensors/transducers
  - psychophysics: determine characteristics
- relative judgements: strong
- absolute judgements: weak
- different optimizations than most machines
  - eyes are not cameras
  - perceptual dimensions not nD array
  - (brains are not hard disks)



## Psychophysical Measurement

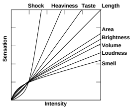
- JND: just noticeable difference
- increment where human detects change
- average to create "subjective" scale
- low-level perception more uniform than high-level cognition across subjects



## Nonlinear Perception of Magnitudes

sensory modalities not equally discriminable

**Stevens' Power Law:** $I = S^p$

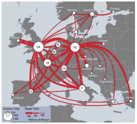



[Stevens, On the Theory of Scales of Measurement, Science 103:2684, 1946]



## Dimensional Dynamic Range

- linewidth: limited discriminability



[mappa.mundi.net/maps/maps_014/telegeography.html]

## Dimensional Ranking: Accuracy

- spatial position best for all types

| Quantitative | Ordinal | Nominal |
|---|---|---|
| Position | Position | Position |
| Length | Density | Hue |
| Angle | Saturation | Texture |
| Slope | Hue | Connection |
| Area | Texture | Containment |
| Volume | Connection | Density |
| Density | Containment | Saturation |
| Saturation | Length | Shape |
| Hue | Angle | Length |
| Texture | Slope | Angle |
| Connection | Area | Slope |
| Containment | Volume | Area |
| Shape | Shape | Volume |

[Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986]



## Cleveland: Perceptual Studies

Cleveland and McGill. Graphical Perception: Theory, Experimentation and the Application to the Development of Graphical Models. J. Am. Stat. Assoc. 79:387, pp. 531-554, 1984.

| Mackinlay | Cleveland (quant) |
|---|---|
| position | position along common scale<br>position along nonaligned scales |
| length<br>angle<br>slope | length, direction, angle |
| area<br>volume<br>density<br>saturation | area<br>volume, curvature<br>shading, color saturation |
| hue | |
| texture | |
| connection | |
| containment | |
| shape | |

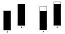


## Weber's Law

- ratio of increment threshold to background intensity is constant
  - relative judgements within modality

$$\frac{\Delta I}{I} = K$$

- Cleveland example: frame increases accuracy

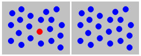


Graphical Perception: Theory, Experimentation and the Application to the Development of Graphical Models. William S. Cleveland, Robert McGill, J. Am. Stat. Assoc. 79:387, pp. 531-554, 1984.



## Preattentive Visual Dimensions

- color (hue) alone: preattentive
  - attentional system not invoked
  - search speed independent of distractor count



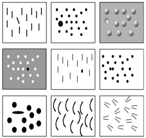
- demo

[Chris Healey, Preattentive Processing, www.csc.ncsu.edu/faculty/healey/PP/PP.html]



## Many Preattentive Visual Dimensions

hue
shape
texture
length
width
size
orientation
curvature
intersection
intensity
flicker
direction of motion
stereoscopic depth
light direction, . . .



[www.csc.ncsu.edu/faculty/healey/PP/PP.html]



## Not All Dimensions Preattentive

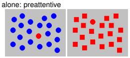
parallelism

[www.csc.ncsu.edu/faculty/healey/PP/PP.html]

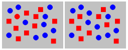


## Preattentive Visual Dimensions

- color alone: preattentive
- shape alone: preattentive



- combined hue and shape (demo)
  - requires attention
  - search speed linear with distractor count



[www.csc.ncsu.edu/faculty/healey/PP/PP.html]



## Separable vs. Integral Dimensions

- only some dimensions separable

color location | color motion | color shape | size orientation | x-size y-size | red-green yellow-blue

[Colin Ware, Information Visualization: Perception for Design. Morgan Kaufmann 1999.]



## Perception: Readings

On the Theory of Scales of Measurement. S. S. Stevens. Science 103:2684, 1946

Graphical Perception: Theory, Experimentation and the Application to the Development of Graphical Models William S. Cleveland, Robert McGill, J. Am. Stat. Assoc. 79:387, pp. 531-554, 1984.

Information Visualization: Perception for Design. Chapter 5: Visual Attention and Information That Pops Out. Colin Ware. Morgan Kaufmann, 2004 (2nd edition).

Perception in Visualization. Christopher G. Healey. http://www.csc.ncsu.edu/faculty/healey/PP/